

単語音声の自動認識に関する研究-言語構造の利用を中心として-

著者	牧野 正三
号	460
発行年	1974
URL	http://hdl.handle.net/10097/9196

氏 名	まきのしょうぞう 牧 野 正 三
授 与 学 位	工 学 博 士
学位授与年月日	昭和 4 9 年 1 0 月 9 日
学位授与の根拠法規	学位規則第 5 条第 1 項
研究科, 専攻の名称	東北大学大学院工学研究科 (博士課程) 電気及通信工学専攻
学 位 論 文 題 目	単語音声の自動認識に関する研究 — 言語構造の利用を中心として —
指 導 教 官	東北大学教授 城戸 健一
論 文 審 査 委 員	東北大学教授 城戸 健一 東北大学教授 大泉 充郎 東北大学教授 野口 正一 東北大学助教授 鈴木 久喜 東北大学助教授 比企 静雄

論 文 内 容 要 旨

第 1 章 緒 論

音声認識の研究は、人間と機械との情報の交換手段として音声の使用を可能にするためのものである。しかし、従来の音声認識の研究は音声波の短時間の物理的な面にのみ依存した認識を行っていたので、調音結合等のため満足すべき成果は得られなかった。一方人間は時間的変化を利用して調音結合を正規化し、更に言語情報を利用して誤り訂正を行なっている。音声の自動認識ではこの言語情報の利用が今迄は考慮されていなかった。本研究は音声認識における言語情報の利用に関する研究である。言語情報を利用すれば特徴抽出がある程度悪くても高い認識率の音声認識システムを経済的に構成する事が可能になる。言語情報である単語辞書を単語音声認識システムに利用する事の有効性は既に報告されているが、これは幾つかの理想化された条件の場合

であり、実際のシステムではこれらの条件はほとんど満たされていない。

本章では単語辞書を利用した単語音声認識システムを実際に構成し、その際問題となったセグメンテーションエラーに対する幾つかの対策を論ずる。

第2章 単語辞書を利用した単語音声の認識

本章では図1、図2に示す単語辞書を利用した単語音声システムを実際に構成し、その問題点を検討した。このような言語情報を利用した単語音声認識システムが実際に構成されたのは初めてのことである。

これら二つのシステムの音声分析と特徴抽出の部分は共通である。音声分析としてはバンドパスフィルターを用いた。パラメータとしては、帯域電力の相互比等を用いた。これらのパラメータを弁別的特徴の性質を利用して線型変換することによって今迄行なわれていなかった弁別的特徴と物理量との対応づけを行ない、平均誤り率10.5%という良好な抽出率を得た。

図1のシステムでは、特徴フレーム間のユークリッド距離 dt と区間の持続時間を用いて一次セグメンテーションを行なった。更に実験的に求めた音素結合規則を用いてセグメンテーションの誤りを訂正し、音節構成の規則性を利用して単語辞書を参照した。男性1名の発声した53単語について認識実験を行なった結果、音素結合規則を用いることによって音素認識率が8.7%向上し、更に単語辞書を用いることによって23.6%向上した。又単語認識率は単語辞書を用いる前には0であったものが単語辞書を用いた後は79.2%と改善された。以上の結果、音素結合規則と単語辞書の利用が実際のシステムにおいても有効であることが確かめられた。

図2のシステムは単語辞書からの合成という考えに基づいてセグメンテーション認識を行なうのであり、合成規則にかわるものとして持続時間辞書の利用を提案し、有効であることを確かめた。更に動的計画法を利用して時間軸整合することによって図1のシステムの実験に用いた資料に対して94.3%、他の男性9名の発声した各10単語について90%という非常に良い単語認識率が得られた。図2のシステムでは一々セグメンテーションをやり直すので時間がかかる。それ故究極の音声認識システムでは図1のシステムで認識できない場合に図2のシステムで認識する事が考えられる。そこで図1のシステムの改善を目標とした。即ち、音素結合規則の自動構成とセグメンテーションエラーに強い単語辞書の利用方法の確立である。

第3章 推移確率の利用による音素系列の訂正

本章では音素結合規則の自動構成を目標とした。音素結合規則は音素間の推移確率の利用方法の一つである。そこでまず任意の単語集合とConfusion Matrixが与えられた場合、セグメンテーションエラーを含んだ音素系列を音素間の推移確率の利用によって訂正する一般的な方法を提案した。このような方法は今迄行なわれていない。更に動的計画法の利用によって計算が簡単

化されることを示した。この方法を図1のシステムの間中間結果に適用した所、最大10.2%の音素群の認識率の改善が得られた。次にこの方法を利用して音素結合規則を自動的に構成する方法を示した。上記と同じく図1のシステムの間中間結果の Confusion Matrix と推移表を利用して音素結合規則を求め、その音素結合規則を改めて中間結果に適用した。その結果最大8.5%音素群の認識率が改善された。音素結合規則はハード化が容易であり、かつ計算時間が極めて少なくなるという利点がある。ここで述べた方法はいずれも任意の単語集合と Confusion Matrix が与えられれば自動的に計算できる一般性の高い方法である。

第4章 音素群系列からの順序特徴行列による単語認識

本章ではセグメンテーションエラーがありかつ音素群しか認識できない入力からでも単語辞書の利用によって単語を認識する方法として順序特徴行列を提案し、その有効性を確かめた。順序特徴行列は任意の単語集合の中から音素の順序関係や音素群の情報だけを用いて、ある単語を規定する特徴行列である。例えば図3Aに示す行が特徴、列が音素位置をあらわす特徴行列で表現された5つの単語 $W_1 \sim W_5$ の中から W_1 を規定するにはBに示す様な特徴行列がわかれば良い。例えば F_1 は〔二〕という特徴行列で行わされる音素が入力のどの位置でも良いから存在すれば W_1 である事を意味し、 F_2 は第2特徴がマイナスである音素が入力のどこかに存在し、更にその後のどこかに第2特徴がマイナスの音素がもう一つ存在すれば W_1 である事を示す。このような特徴行列を順序特徴行列と呼ぶ。この順序特徴行列を用いる事によってセグメンテーションエラーのある音素群系列から単語を認識できる。166単語を単語辞書の語彙とし、その中の6~8音素単語113単語に種々の誤りを生起させて順序特徴行列の訂正能力を評価した所、1音素程度のセグメンテーションの誤りでも約90%の訂正率があった。この方法を図1のシステムの間中間結果に適用した所84.9%と5.7%単語認識率が改善された。更に第3章の結果に適用した所86.8%と6.6%単語認識率が改善された。この方法はハード化が容易であり、一部の音素群の認識が終った段階で単語認識ができるという利点がある。

第5章 順序特徴行列で構成された単語辞書の構成と検索

本章では、順序特徴行列を用いる場合の欠点である項目数の増加による使用記憶領域と検索時間の増大による影響を減少させるために、順序特徴行列を用いた場合の辞書の構成と検索にデータ構造の表検索の方法を利用する事を提案した。まず表検索の方法のうちハッシュ検索法とランクパス検索法を利用する場合の基本的な方法を示した。次にこれらの方法の改善を行なった。ハッシュ検索法では、各音素に共通する音素群特徴を用いてハッシュ変換する事によって検索時間を減少させる事を提案した。又ハッシュ変換によって項目が登録されている行と、項目が登録されず、

コンフリクト処理のために使う行とを区別するチェックビットを設ける事を提案した。これらの改善方法によって検索時間が減少する事を理論的に示した。一方ランクパス検索では特徴の順序性を利用して検索する方法を提案した。この方法が表検索の方法を用いない場合に較べて必ずしも検索時間が減少する事を理論的に示した。これらの方法を実際にプログラミングし、第4章の結果を用いてシミュレーションを行なった。その結果表検索の方法を用いない場合に較べて、ハッシュ検索法の改良法で $1/5$ 、ランクパス検索法の改良法で $1/11$ に検索時間が減少した。又ランクパス検索では記憶領域も減少させる事を明らかにした。以上の結果、ランクパス検索法の改良法で順序特徴行列で構成された単語辞書の検索に用いるのが最も有効である事が明らかになった。

第6章 結 論

本研究は、言語情報を利用する際の問題点であるセグメンテーションエラーについて考察研究し、孤立単語音声の認識に利用できる二つの言語情報、即ち、単語辞書と音素間の推移確率を音素の誤認識とセグメンテーションエラーのある音素系列に適用する方法を確立し、その有効性を明らかにした。又始めて言語情報を利用した単語音声認識システムを構成した。

本研究を行なうにあたり終始御指導戴いた東北大学城戸健一教授、鈴木久喜助教授、板橋秀一前助手に深く感謝致します。

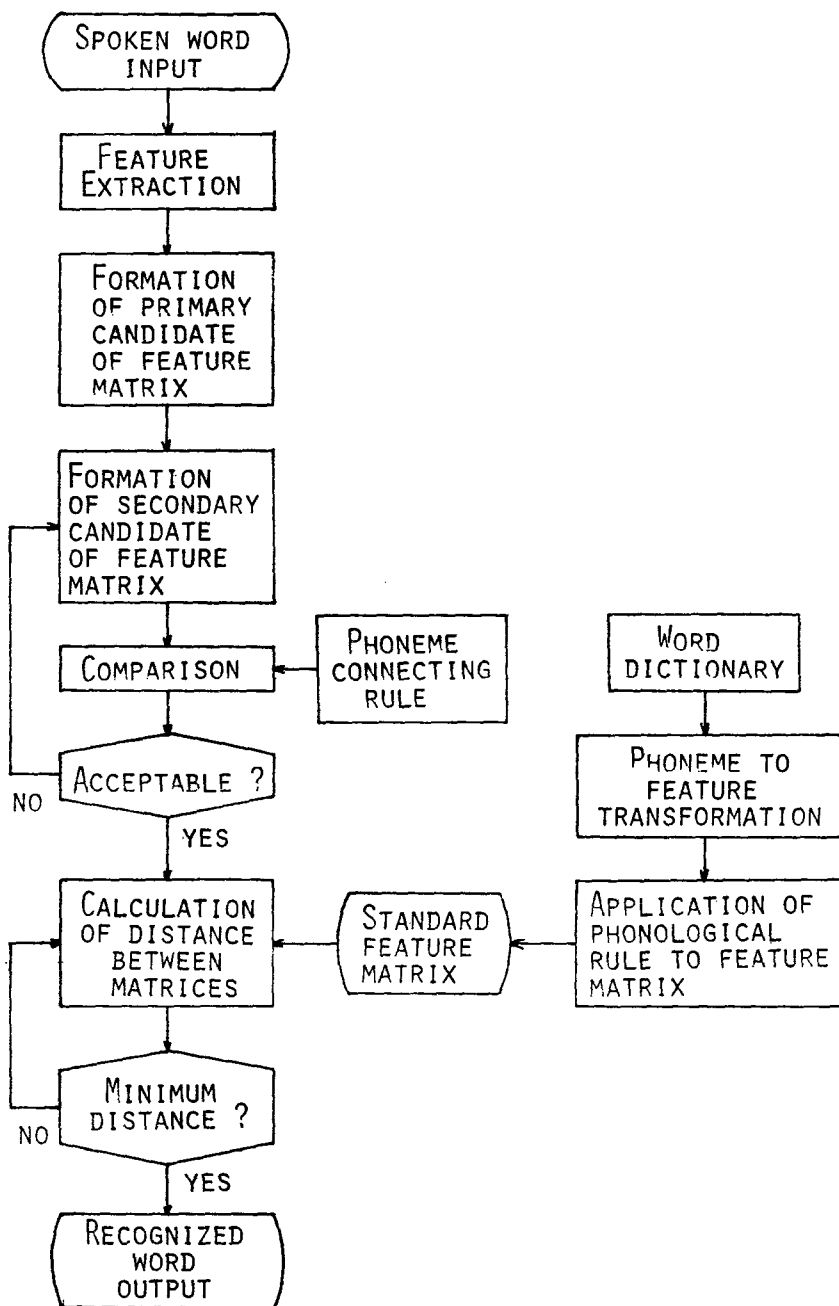


Fig. 1 Schematic diagram of spoken word recognition system (1) using word dictionary and phonological rule.

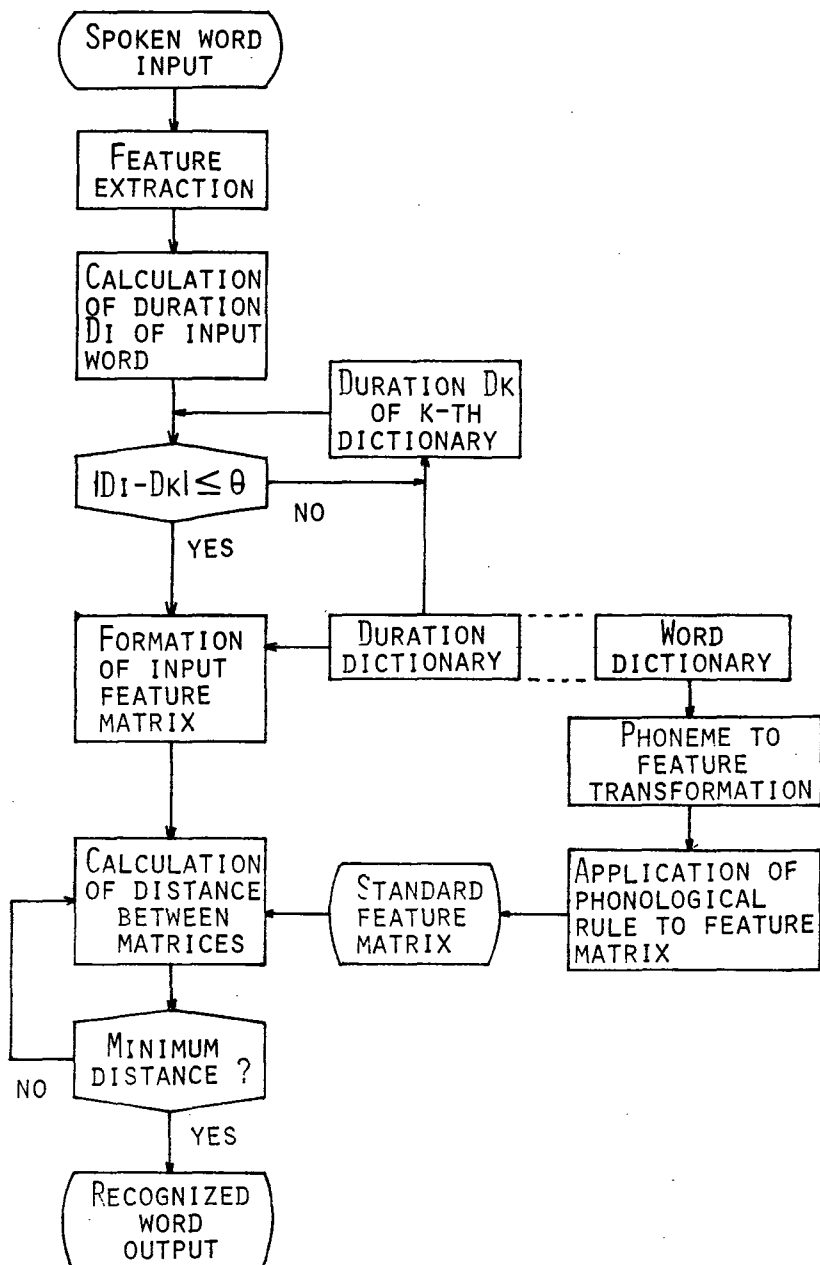


Fig. 2 Schematic diagram of spoken word recognition system (2) using word dictionary and duration dictionary.

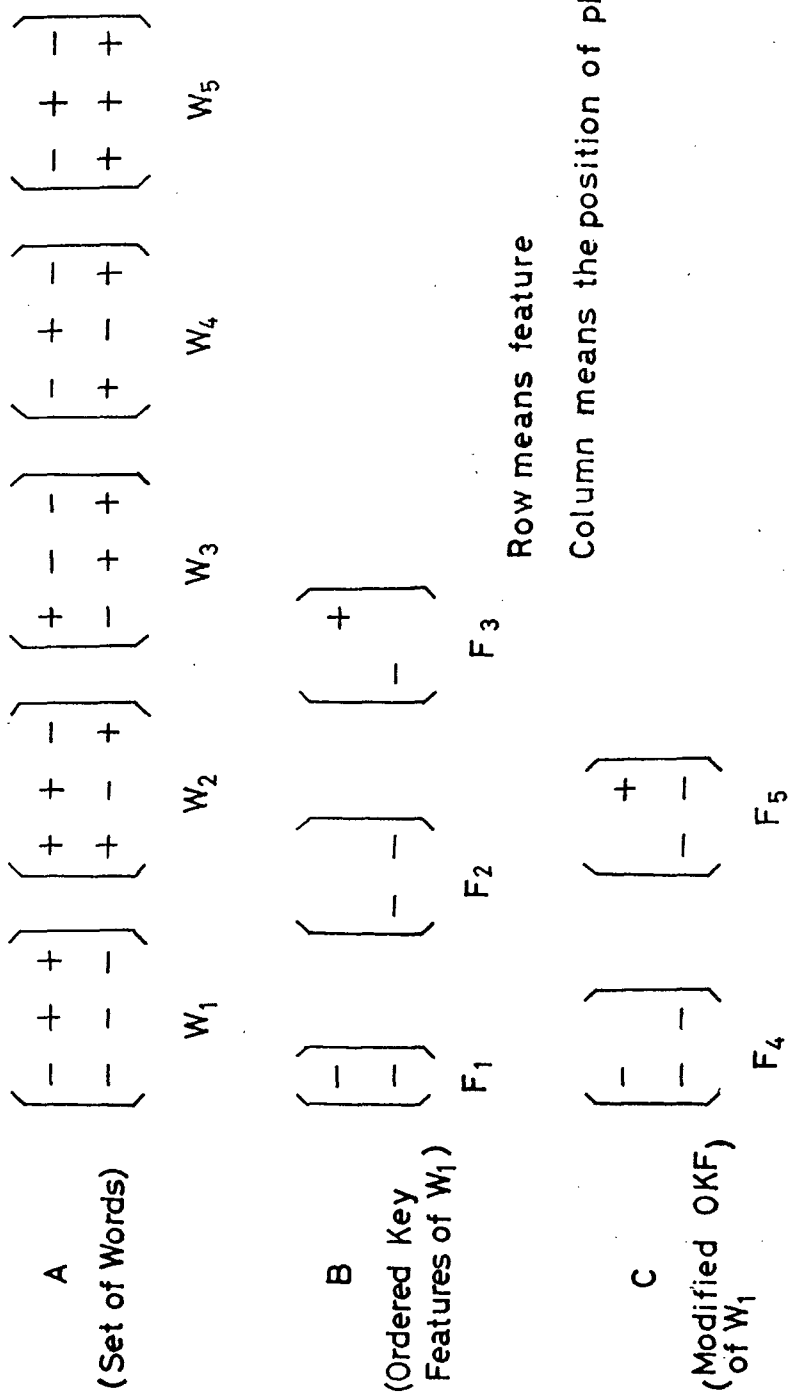


Fig. 3 Ordered Key Features and Modified OKF's extracted from word set ($W_1 \sim W_5$)

審 査 結 果 の 要 旨

言語音声は、人間の情報出力の手段として、手軽で、かつ、最も高速度のものであるため、人間から機械への情報の伝送を言語音声で行うことができれば、その利益は大きいと考えられている。そのため、音声自動認識の研究は古くから行われてきた。しかし、音声は発声者ごとにその音響的性質が異なるばかりでなく、前後の言葉の違いによっても音響的性質が変わり、その部分だけを切り取って聴取すると、他の発音の音声と混同することが多いというような複雑な様相を呈するので、音響的性質の解明を中心として進められた従来の研究には、疑問が投ぜられている。

本論文の著者は、認識対象を限定語いの単語音声に限ることと、単語の言語構造を利用することによって、音声自動認識を可能とする道を開くことを意図して研究を進めた。本論文はその成果をまとめたもので、6章からなる。

第1章は緒論で、本研究の意義と基本的な考え方を論じている。

第2章では単語辞書を利用した単語音声認識の方法論を、著者の行った研究を中心として論じ本研究の方向づけと重点とを明らかにしてある。

第3章では、音響分析の結果から得られた音素系列を、音素間の推移確率を利用して訂正することを提案し、実用的な方法を与えると共に、その有効性を事例によって示している。本方法は音素の誤認識のみならず、脱落、付加に対しても訂正能力をもつことが特徴である。

第4章では、著者の発案になる順序特徴行列の利用について述べている。ここでは、順序特徴行列を定義すると共に、その誤り訂正能力を検討し、かつ、単語音声の認識実験によってその有効性を確かめている。順序特徴行列は、これを利用することにより、多くの場合に、単語を構成する音素が全部認識されなくても単語が認識できる有用なものである。

第5章は、順序特徴行列で構成された単語辞書の構成と検索について述べたもので、データ構造の利用と辞書検索法の改良とにより、辞書索引に要する時間を短縮できることを示し、本論文の方法の実用性を裏付けている。

第6章は結論である。

以上要するに、本論文は、単語音声の自動認識に単語の言語構造を利用することを中心として研究し、単語音声自動認識の具体的な方法について新たな知見を加えたもので、音声情報処理並びに通信工学の進歩に寄与するところ少なくない。

よって、本論文は工学博士の学位論文として合格と認める。